

Kathrin Steyer / Wolfgang Teubert

DEUTSCH-FRANZÖSISCHE ÜBERSETZUNGSPLATTFORM **Ansätze, Methoden, empirische Möglichkeiten**

Abstract

Der Artikel diskutiert Ziele, Methoden und Probleme einer geplanten deutsch-französischen Übersetzungsplattform. Auf der Basis paralleler und vergleichbarer Korpora sollen mit Hilfe dieses elektronischen Werkzeuges nicht nur Übersetzungsvorschläge für Einzelwörter, sondern auch für Kollokationen, Phrasen und systematisierte Verwendungskontexte gemacht werden. Dabei geht es vor allem um die Erfassung jener Einheiten, die nicht in traditionellen Wörterbüchern stehen, aber bereits Usus sind. Das Projekt integriert drei Herangehensweisen: Korpusbasiertheit, Orientierung auf Idiomaticität als relevantes Übersetzungsprinzip, Kontextbezogenheit. Der Beitrag umreißt den Projektansatz anhand der Kollokationsproblematik.

This article discusses the aims, methods and problems of a planned German-French translation platform. On the basis of parallel and comparable corpora, this electronic tool is intended to give suggested translations not only for individual words, but also for collocations, phrases and classified contexts. The main aim is to incorporate units which are not listed in traditional dictionaries, but which have become commonly used. The project incorporates three approaches: it is corpus-based, it is oriented towards idiomaticity as a relevant translation principle, and it is context-related. The article describes the concept of the project and provides examples to illustrate the problems in the field of collocations.

0. Vorbemerkung

In den folgenden Ausführungen sollen Ziele, Methoden und damit verbundene Probleme der einzelnen Phasen des Langzeitvorhabens „Deutsch-Französische Übersetzungsplattform“ voneinander abgegrenzt, näher erläutert und – was das erste unmittelbare Ziel betrifft – konkret beschrieben werden.

1. IDS-Langzeitvorhaben „Deutsch-Französische Übersetzungsplattform“

Eine Konzeption dieser Größenordnung muß sich einigen grundsätzlichen Fragen stellen, die wie folgt zusammengefaßt werden können:

- Inwieweit geht dieses lexikographische Vorhaben qualitativ über die bisherigen großen Wörterbuchunternehmen hinaus?
- Welche neue Qualität bietet es sowohl für sprachwissenschaftliche Erkenntnisinteressen als auch für die praktische Übersetzungsarbeit?
- Lassen sich anwendbare Lösungen entwickeln, die dem Übersetzer ein neuartiges, dynamisches und flexibles Werkzeug in die Hand geben? Welche Prozesse können im Projektverlauf automatisiert werden, und an welcher Stelle ist die „ordnende Hand“ des Linguisten unabdingbar?
- Welche Ergebnisse kann ein solches Projekt sowohl für linguistische Forschungen in der jeweiligen Einzelsprache als auch für kontrastive Analysen erbringen?

Auf einige dieser Punkte gehen wir im folgenden ein.

Die geplante Übersetzungsplattform gehört sowohl in sprachtechnologischer als auch linguistischer Hinsicht zu einer neuen Generation lexikographischer Werkzeuge. Sie soll vor allem Nutzern mit durchschnittlichen Sprachkenntnissen helfen, sich in der jeweils fremden Sprache selbständig und angemessen auszudrücken bzw. Texte in der Fremdsprache adäquat zu formulieren. Teubert faßt die wichtigsten Leistungen folgendermaßen zusammen:

- „Es soll Übersetzungsvorschläge nicht nur für das Einzelwort, sondern für ganze Phrasen [...] und für Kollokationen anbieten.
- Es soll, was den Allgemeinwortschatz von 20.000 bis 30.000 Wörtern angeht, das syntagmatische und paradigmatische Verhalten der lexikalischen Einheiten möglichst vollständig in Hinsicht auf die Benutzerbedürfnisse erschließen [...].
- Es soll das, was bereits übersetzt ist, abspeichern und wieder anbieten, wenn die Übersetzungseinheit in gleicher oder vergleichbarer Form wiederkehrt.“ (Teubert 1996b, S. 9)

Er sieht dabei vor allem zwei Vorteile, die elektronische Übersetzungswerkzeuge gegenüber herkömmlichen – gedruckten – haben werden: *Vollständigkeit* und *Zuverlässigkeit*.

Der Projektidee liegen Einsichten zugrunde, die sowohl von an Sprachverwendung interessierten Lexikologen, Semantikern und Lexikographen, von Sozio- und Textlinguisten einerseits als auch von Computer- und Korpuslinguisten andererseits zur Zeit mehr und mehr ins Zentrum der wissenschaftlichen Diskussion gerückt werden. Immer deutlicher treten dabei die Desiderate bisheriger linguistischer Analyse und das in weiten Teilen noch recht beziehungslose Nebeneinander vieler der erwähnten Forschungsdisziplinen zutage. Die neue Qualität der geplanten Übersetzungsplattform wird darin bestehen, verschiedene Herangehensweisen zu integrieren und eine Synthese folgender Aspekte vorzunehmen: a) *Korpusbasiertheit*, b) *Orientierung auf Idiomatizität als ein relevantes Übersetzungsprinzip* und c) *Kontextbezogenheit*. Für alle diese Herangehensweisen und Paradigmen gibt es bereits computerlinguistische und auch lexikographische Erklärungs- bzw. Lösungsversuche. Kein Vorhaben hat sie jedoch vollständig zusammengeführt.

a) Korpusbasiertheit

Eine Sprache zu erforschen, ohne die Resultate anhand der tatsächlichen Sprachverwendung zu verifizieren, erscheint nur für ganz bestimmte theoretische Erkenntnisziele noch sinnvoll. Empirische Fundierung muß auf der Basis großer, anwendungsorientiert ausgewogener Korpora erfolgen. Gerade in sprachtechnologischer Hinsicht hat sich eine Entwicklung vollzogen, die von den Linguisten bis vor kurzem noch nicht ausreichend zur Kenntnis genommen wurde. In besonderem Maße betreffen diese Neuorientierungen den Bereich der Lexikographie, vor allem der bi- und multilingualen Lexikographie, wo ein korpusbasiertes Herangehen zunehmend akzeptiert, gefordert und auch schon praktiziert wird.¹

¹ Vgl. Boguraev/Pustejovskys Plädoyer für eine Korpuslexikographie (1996, S. 10ff.).

Das verdeutlicht genau ein Problem der bisherigen zweisprachigen Lexikographie, die unbestritten bedeutende Leistungen hervorgebracht hat: Die Entscheidung für Wörterbucheinträge ist nach wie vor zu einem großen Prozentsatz von der intuitiven Bewertung der Lexikographen abhängig. Sie entscheiden auf Grund ihrer Sprachkompetenz, ob sie eine Variante für einschlägig und relevant halten. Auch bei jenen lexikographischen Projekten, die bereits auf großen Korpora aufbauen, wird schließlich eingeräumt, diese in letzter Instanz doch nur zur Validierung selbst ausgewählter Belege genutzt zu haben. Die eigene Sprachkompetenz und -erfahrung bleibt Maßstab für Relevanzsetzungen und für im Wörterbuch zu fixierende Kontextangebote (vgl. u.a. Zimmer 1990, Schemann 1994). Es wurden keine automatisierten statistischen Frequenzanalysen durchgeführt, die zu einer ungleich größeren, manuell nicht mehr zu bewältigenden Belegmenge führen.

Mit der konsequenten Einbeziehung von maschinenlesbaren Korpora und dem lexikographischen Produkt in Form 'elektronischer Datenbanken' sollen diese intuitiven Zuordnungen zugunsten einer Erforschung dessen, was Usus ist, überwunden werden. Es geht darum, wirklich zentrale Verwendungen auf der Basis von quantitativen Kriterien (*statistische Signifikanz*) und qualitativen Kriterien (*Systematisierung von Sprachverwendungskontexten*) lexikographisch zu kodifizieren und abrufbar zu machen.

Der auch in der Korpuslinguistik noch relativ neue Ansatz, zwei- oder mehrsprachige Übersetzungswerkzeuge auf der Basis *paralleler und vergleichbarer Korpora*² zu entwickeln, ist vor allem durch die eben dargestellten Hintergründe motiviert. Parallelkorpora bestehen aus Originaltexten und deren Übersetzungen bzw. äquivalenten Übersetzungstexten in verschiedenen Sprachen. Ein bereits existierendes Korpus ist das „Canadian Hansard Corpus“, das die Sitzungsprotokolle des kanadischen Parlaments jeweils auf englisch und französisch enthält. Alle Texte werden durch automatische Verfahren miteinander in Beziehung gesetzt und können bis in ihre kleinsten Elemente hinein miteinander verglichen werden (vgl. dazu 2.). Der Vorteil dieser Parallelkorpora besteht darin, daß sie eine Vielzahl von Übersetzungsvarianten bieten, die weder ein einzelner Lexikograph noch ein einzelner Übersetzer in dieser Menge abrufen kann. Sie enthalten das implizite Sprachwissen des Übersetzers, das nicht nur mehr umfaßt, sondern anders organisiert ist als die expliziten Informationen, die in bilingualen Wörterbüchern enthalten sind. Dieses implizite Übersetzungswissen muß durch geeignete Verfahren verfügbar gemacht werden. So läßt sich beispielsweise mit statistischen Frequenzanalysen ermitteln, welche Übersetzungsvarianten wirklich häufig eingesetzt werden und welche Kombinationsmöglichkeiten von sprachlichen Einheiten quantitativ bedeutsam sind, aber nicht im Wörterbuch vorkommen.

Da bei solchen parallelen Korpora naturgemäß ein asymmetrisches Verhältnis zwischen den Ausgangstexten und ihren Übersetzungen besteht, wird vorgeschlagen, darüber hinaus vergleichbare Korpora zur Ergebnisvalidierung zu nutzen. Vergleichbare Korpora enthalten Texte, die in Größe, Aufbau (nach sprachlichen und externen Merkmalen) gleich strukturiert, aber keine Übersetzungen voneinander sind.

² Zur Idee paralleler und vergleichbarer Korpora vgl. ausführlich Teubert (1996a).

b) Orientierung auf Idiomatizität als relevantes Übersetzungsprinzip

Das traditionelle, gedruckte zweisprachige Wörterbuch kann nicht anders als auf das alphabetisch geordnete Einzelwort bezogen sein. Je mehr sich moderne Lexikographen bemühen, die Idiomatik zu berücksichtigen, desto unübersichtlicher müssen notgedrungen die Worteinträge sein. Übersetzt wird aber selten Wort für Wort, sondern in der Regel – mit Ausnahmen von strikt terminologisch bestimmten Übersetzungen – in größeren Übersetzungseinheiten: Mehrwortkonstruktionen, Kollokationen, idiomatischen Wendungen usw. Hausmann bezeichnet diese sprachlichen Gebilde als „Halbfertigprodukte der Sprache“ (vgl. Hausmann 1985, S. 118). Gerade auch durch die Forschungen auf dem Gebiet der automatischen Übersetzungen ist deutlich geworden, daß Übersetzungsphänomene und -probleme oft in direktem Zusammenhang mit solchen Kombinationsmustern einer Sprache stehen, das heißt, daß Übersetzungsvorgänge in starkem Maße über die Aktualisierung von idiomatischen Wendungen ablaufen.

„[...] die Kenntnis der Kollokationen ist nicht etwa sprachlicher Luxus, sondern elementar. [...] In den Kollokationen ist in besonderer Weise die unverwechselbare Eigenart einer Sprache niedergelegt.“ (Hausmann 1993, S. 9)³

Für ein neuartiges Übersetzungswerkzeug kann dies nur bedeuten, solche mehr oder weniger festen Wortverbindungen in den Mittelpunkt des Interesses zu rücken, sie operationalisierbar zu machen und effektive Lösungen zu finden.

c) Kontextabhängigkeit

Die geplante Übersetzungsplattform soll kontextabhängige Übersetzungsäquivalente für Kollokationen und ganze Phrasen anbieten. Der Begriff 'Kontext' spielt in diesem Konzept eine fundamentale Rolle. Sinclair verweist erneut darauf, daß der Terminus 'Kontext' in der Linguistik in zweierlei Weise verwendet wird: Zum einen bildet er die unmittelbare sprachliche Umgebung von Ausdrücken ab (also die unmittelbaren syntaktischen und lexikalisch-semantischen bzw. textuellen Strukturen). Zum anderen wird mit 'Kontext' auch die situative Einbettung, der soziokulturelle background erfaßt (vgl. Sinclair 1991, S. 171). Mit einer systematischen Analyse der jeweiligen *sprachlichen Umgebungen* für Übersetzungseinheiten wird man möglicherweise dann auch zu Aussagen über kommunikativ-situative Kontexte idiomatischer Wendungen (z.B. Textsortenspezifik, Stilebene etc.) kommen können. Dabei ist bisher zuwenig beachtet worden, daß bei der Wahl eines bestimmten Übersetzungsäquivalents in der Zielsprache auch die Kontextrestriktionen in der Zielsprache zu suchen und nicht allein die kontextuellen Einbettungen der Übersetzungseinheit in der Ausgangssprache zu berücksichtigen sind. Die Bedeutungsaufgliederung eines Wortes in einem monolingualen Wörterbuch ist für die Auswahl des passenden Übersetzungsäquivalents gerade in schwierigen Fällen kaum hilfreich.

Die geplante Übersetzungsplattform ist mehr als ein gedrucktes oder elektronisches Wörterbuch. In ihrem modularen Aufbau kombiniert sie lexikalische Da-

³ Zu Idiomatikansätzen und entsprechenden Forschungsperspektiven vgl. ausführlich Sandig (1994).

tenbank von Übersetzungsäquivalenten, Translation memory und die auf die konkrete Anwendung (den Übersetzungsfall) bezogene Auswertung des Parallelkorpus, so daß sich der Übersetzer auf die Wahl zwischen angebotenen Alternativen und die Übersetzung neologischer Elemente konzentrieren kann. Dieses Langzeitvorhaben kann nur im Verbund mit anderen, durch gemeinsame Interessen verbundenen Sprachinstituten in Europa verwirklicht werden.

2. Korpusbasiertes und kontextorientiertes Kollokationswörterbuch

Wir beabsichtigen, als erstes lexikographisches Ergebnis ein zweisprachiges Kollokationsverzeichnis in gedruckter und elektronischer Form zu erarbeiten, das im oben angeführten Sinne korpusbasiert und kontextorientiert ist.⁴ Mit der lexikographischen Datenbank soll darüber hinaus die Interdependenz statistik-, listen- und regelbasierter Ansätze untersucht werden. Dabei wird zu zeigen sein, wie man parallele und vergleichbare Korpora für die mehrsprachige Lexikographie und für kontrastive sprachwissenschaftliche Analyse nutzbar machen kann. Ein weiteres Resultat stellt die Erarbeitung einer tiefergehenden Systematik für Kollokationen generell dar, die die Grundlage für weitere Forschungen bilden kann.

2.1 'Kollokation' – ein Arbeitsbegriff

Unter *Kollokationen* verstehen wir mit Sinclair das Vorkommen zweier oder mehrerer Wörter innerhalb eines bestimmten Abstandes im Text, wobei wir uns bei einem ersten Herangehen auf signifikante und relevante Zweierkombinationen konzentrieren werden.⁵

„Collocation is the occurrence of two or more words within a short space of each other in a text.“ (Sinclair 1991, S. 170)

Wesentliche Kategorien sind dabei 'node', 'collocate' und 'span'.⁶

Hausmann versteht den Terminus 'Kollokation' vor allem als charakteristische Kombination zweier Wörter:

„On appellera collocation la combinaison caractéristique de deux mots dans une des structures suivantes:

- a) substantif+adjectif (épithète)
- b) substantif+verbe
- c) verbe+substantif (objet)

⁴ Es gibt bereits einige Kollokations- bzw. Kontextwörterbücher für das Sprachenpaar Deutsch-Französisch/Französisch-Deutsch. Verwiesen sei hier vor allem auf Zimmer (1990); Bárdosi/Ettinger/Stölting (1992); Ilgenfritz/Stephan-Gabinel/Schneider (1993); Schemann (1994).

⁵ Eine ausführliche Diskussion des britischen Kontextualismus, der grundlegende Forschungen zum Problemfeld lexikographischer Erschließung und Auswertung von Korpora geleistet hat, findet in sich bei Lehr (1996) und Bahns (1996), zu einigen Aspekten auch Hausmann (1985).

⁶ Sinclair definiert diese Kategorien folgendermaßen: 'node' „The node word in a collocation (q.v.) is the one whose lexical behaviour is under examination.“ (1991, S. 175); 'collocate' „A word which occurs in close proximity to a word under investigation is called a collocate of it.“ (ebd., 170); 'span' „This is the measurement, in words, of the co-text (q.v.) of a word selected for study. A span of -4,+4 means that four words on either side of the node word (q.v.) will be taken to its relevant verbal environment.“ (ebd., S. 175)

- d) verbe+adverbe
- e) adjectif+adverbe
- f) substantif+(prép.)+substantif" (Hausmann 1989, S. 1010)

Wichtig erscheint uns dabei, Kollokationen als *charakteristische* Wortverbindungen aufzufassen. Das Ziel unserer Arbeit besteht darin, u.a. durch Frequenzanalysen aus einer Vielzahl von in den Texten vorkommenden Wortkombinationen die charakteristischen und typischen herauszufinden. Es interessieren also Kollokationen in ihrer tatsächlichen Verwendung. Dabei soll über die frequentielle Erfassung hinaus letztendlich jener Status von Kollokationen näher beschrieben werden, den Hausmann folgendermaßen charakterisiert:

„La collocation se distingue de la combinaison libre (*the book is useful/das Buch ist nützlich/le livre est utile*) par la combinabilité restreinte (ou affinité) des mots combinés (*feuilleter un livre vs. acheter un livre*). La collocation se distingue d'autre part des locutions (*idioms, Redewendungen, par ex. monter un bateau à qn/jdn. durch den Kakao ziehen/to pull sb's leg*) par son non-figement et par sa transparence." (ebd.)

Dieses Spannungsfeld zwischen völlig freien Wortverbindungen einerseits und festen idiomatischen Redewendungen andererseits läßt sich auf der Basis großer Korpora umfassender analysieren als das bisher geschehen ist.

Das Kollokationswörterbuch soll in mehrfacher Hinsicht über die bereits erwähnten Kontextwörterbücher hinausgehen: Es wird angestrebt, auf der Basis eines bereits vorhandenen großen Parallelkorpus und mit Hilfe vergleichbarer Korpora durch das Zusammenspiel statistisch und linguistisch determinierter Verfahren zentrale Wortverbindungen in den jeweiligen Einzelsprachen in ihrer konkreten Sprachverwendung herauszufinden und sie mit entsprechenden Kollokationen der anderen Sprache zu vergleichen. Als Ergebnis werden sowohl (Frequenz-)Listen von Kollokationen und ihren relevanten Übersetzungsäquivalenten als auch häufig wiederkehrende Kontexte, in die die Kollokationen eingebettet sind, in systematisierter Form angeboten (vgl. dazu 2.4.).

2.2 Das Korpus

Die empirische Basis für das Wörterbuch soll folgendes Parallelkorpus bilden: Seit 1995 erscheint die französische Wochenzeitung „le monde diplomatique“ gleichzeitig in verschiedenen Sprachen, darunter auch in einer deutschen Übersetzung. Es handelt sich um ein exemplarisches Parallelkorpus aus der unmittelbaren Sprach- und Übersetzungspraxis mit ca. 1 Million Wortformen pro Sprache. Um die herausgefundenen Regularitäten zu verifizieren, sollen zusätzlich vergleichbare Korpora in die Analyse einbezogen werden, das heißt vergleichbare Texte von jeweils einer anderen Tages- bzw. Wochenzeitung (z.B. „le monde“ und „Frankfurter Rundschau“).

Mit der Entscheidung für diese Zeitungskorpora wird auch eine Eingrenzung des zu erfassenden Kollokationsvorkommens und der entsprechenden Übersetzungsäquivalente vorgenommen, was für diese Erprobungsphase sinnvoll erscheint. Zeitungskorpora sind in mehrfacher Hinsicht nutzbringend: Zum einen handelt es sich um einen abgegrenzten, gut strukturierten Kommunikationsbereich, der vergleichsweise klar definierte und in beiden Ländern analoge Textsorten (Genres) aufweist. Zum anderen bildet der Bereich der öffentlichen Kommunikation sowohl den politisch-

gesellschaftlichen Diskurs allgemein als auch große Teile der Alltagssprache ab und ermöglicht somit verallgemeinerbare Schlußfolgerungen für den Gebrauch von Kollokationen generell. Auch die internen sprachlichen Strukturen der Texte sprechen für die Auswahl dieses Parallelkorpus. Texte der öffentlichen Kommunikation weisen in besonderem Maße sich wiederholende Textmuster und sprachliche Standardformulierungen auf. Hier vermuten wir deshalb ein großes Potential an signifikanten Kollokationen, festen Wortverbindungen und idiomatischen Wendungen.

Wesentlich erscheint uns schließlich auch ein sprachexterner Grund: Öffentliche Kommunikation gewinnt in Europa zunehmend an Bedeutung. Interessanterweise ist bereits in der Übersetzungswissenschaft darauf hingewiesen worden, daß die Beschreibung von Übersetzungsmechanismen außerhalb literarischer Texte ein Desiderat darstellt. Das trifft in besonderem Maße auf Texte der öffentlichen Kommunikation zu, die bisher – anders als literarische Texte oder Fachtexte – zumeist auf den nationalen Rahmen beschränkt waren (vgl. Stolze 1993). Politik wurde national verhandelt, Medientexte wurden vor allem einsprachig produziert und rezipiert. Die zunehmende Europäisierung der Politik bringt auch ein neues Bedürfnis hervor, nämlich die einschlägigen Texte schnell in andere Sprachen übersetzen zu können.

2.3 Analysemethoden – ein Fallbeispiel

Da wir zum jetzigen Zeitpunkt noch nicht über das eben beschriebene Parallelkorpus in elektronischer Form verfügen, haben wir ein vorhandenes Parallelkorpus aus einem anderen Projektkontext genutzt, um im Rahmen einer Pilotuntersuchung zu testen, inwieweit die zu entwickelnden Methoden einen Erkenntnisgewinn bringen können. Die Analysebasis bildeten die deutsche und die französische Übersetzung von Platons philosophischer Abhandlung „Der Staat”.⁷ Im Zusammenhang mit diesem Beispieltext ergeben sich aber mindestens zwei Probleme: Es handelt sich zum einen um eine sehr spezifische Textsorte (philosophische Abhandlung), aus deren Analyse auf Grund ihrer zum Teil archaischen sprachlichen Strukturen und ihres sehr eingegrenzten Adressatenkreises keine verallgemeinerungswürdigen und repräsentativen Schlußfolgerungen gezogen werden können. Zum anderen ist der Vergleich dadurch erschwert, daß beide Versionen Übersetzungen aus einer dritten Sprache, dem Griechischen, sind. Sie stellen also keine Transformationstexte voneinander dar, ihr Abstand zueinander ist größer als in Parallelkorpora üblich. Daher konnte es nicht um einen Übersetzungsvergleich im eigentlichen Sinne gehen, also um die Frage, welche der beiden Fassungen das Original auf welche Weise abbildet. Möglich war jedoch ein Vergleich zweier äquivalenter Übersetzungen, was einem Perspektivenwechsel gleichkommt: Es interessiert nicht das Verhältnis Original – Wiedergabe mit entsprechenden Bewertungen der Übersetzungsqualität, vielmehr werden zwei Übersetzungsvarianten als selbständige Einzeltexte behandelt, die eine äquivalente Textstruktur aufweisen. Dabei lassen sich identische, aber durchaus auch sehr unterschiedliche sprachliche Realisierungen feststellen, so daß es im folgenden zutreffender ist, von sprachlichen *Parallelvarianten* und nicht von *Übersetzungsäquivalenten* zu sprechen.

⁷ Dieses Parallelkorpus wurde im Rahmen des EG-Verbundprojektes TELRI (Trans-European Language Resources Infrastructure) erstellt und umfaßt mehrere elektronische Übersetzungen.

```

<text>
<body>
<83>
<div type=book>
<head>Erstes Buch</head>
<milestone n=327>
<milestone n=a>
Gestern ging ich mit Aristons Sohn Glaukon
zum Peiraieus hinab, um zur G4ouml;tin zu
beten, und zugleich wollte ich mir den
Festzug ansehen, den man ja jetzt zum
erstenmal veranstaltete. Pr4auml;chtig war
er nun wirklich, der Festzug der Einheimischen,
nicht minder pr4auml;chtig erschien mir auch
der der Thraker.
<milestone n=b>
Da wir gebetet und alles betrachtet hatten,
machten wir uns auf den Heimweg zur Stadt.
Dabei erblickte uns von ferne Polemarchos,
der Sohn des Kephalos, und schickte uns eilends
seinen Diener nach mit der Bitte, auf ihn zu
warten. Der Diener fa4szlig;te mich hinten am
Mantel und sagte: "Polemarchos ersucht euch,
auf ihn zu warten!" Ich wandte mich um und
fragte, wo er denn sei. "Er kommt dort hinten
nach. Wartet doch!" "Dann wollen wir warten",
sprach Glaukon.
<milestone n=c>

```

```

<text>
<body>
<div type=book>
<head>LIVRE I </head>
<milestone n=327>
J'4eacute;tais descendu, hier, au Pir4eacute;
e avec Glaucon fils <lb> d'Ariston
[[onesuperior]]. Mon intention 4eacute;tait
d'adresser un voeu 4agrave; la <lb>
d4eacute;esse[[twosuperior]],et en meme temps
je voulais voir la fete: comment allaient-ils
s'y prendre, vu qu'ils la
c4eacute;l4eacute;braient pour <lb> la
premi4egrave;re fois ? Sans doute, je trouvai
belle la procession <lb> des gens du pays, elle
aussi, mais ne me parut pas moins <lb>
adapt4eacute;e 4agrave; la f4eacirc;te la
fa4ccedil;on dont d4eacute;fil4egrave;rent les
Thraces. Une <lb> fois nos voeux
adress4eacute;s, et le spectacle
regard4eacute;, <milestone n=b>nous <lb>
repartions vers la ville. Or, nous voyant de
loin nous h4acirc;ter <lb> de rentrer chez
nous, Pol4eacute;marque fils de
C4eacute;phale[[threesuperior]] fit <lb> courir
son petit serviteur pour nous inviter 4agrave;
l'attendre. <lb> Le petit, m'attrapant
par-derr4egrave;re par mon manteau: <lb>
4t;4t; Pol4eacute;marque, dit-il, vous invite
4agrave; l'attendre... 4t;4t; Alors je
<lb> me retournai, et lui demandai 4ugrave;
4eacute;tait le ma4icirc;tre lui-m4eacirc;me.
<lb> 4t;4t; Le voici, dit-il, il arrive
derr4egrave;re moi. Allez, attendez. -
<lb> Eh bien nous attendrons 4t;4t;,
dit Glaucon. <lb>

```

Abbildung 1

Polemarchos sagte nun: "Sokrates, ihr seid wohl alle auf dem Heimweg zur Stadt?"

"Nicht 4bel, dein Scharfsinn!" antwortete ich.

"Du siehst wohl unsere 4berzahl?"

"Nat4rlich!"

"Entweder nehmt ihr es mit uns auf - oder ihr m4sst dableiben!"

Pol4eacute;marque dit alors: - Socrate, on dirait que vous vous h4tez de rentrer en ville.

- Ta conjecture n'est pas fausse, dis-je.

- Et vois-tu combien nous sommes ? dit-il.

- Impossible de ne pas le voir !

- Alors, dit-il, ou bien montrez-vous plus forts que les hommes que voici; ou bien restez ici.

Abbildung 2

Anhand dieser Fallstudie sollten erste Methoden experimentell erprobt werden, sowohl was die computertechnische Seite als auch was das linguistische Erklärungspotential betrifft. Hinsichtlich der computertechnischen Seite konnten wir auf Erkenntnissen der EG-Korpusprojekte TELRI und PAROLE aufbauen, die von W. Teubert für das IDS koordiniert werden. Ein Kernpunkt ist dabei das Problem des 'lexical alignment' in Parallelkorpora, also die Zuordnung von Übersetzungsäquivalenten in parallelen Texten, wozu zur Zeit Verfahren entwickelt werden. Es geht dabei um die automatische und optisch sichtbar gemachte Parallelisierung der Textkorpora, in unserem Fall also der deutschen und der französischen Übersetzung von Platos „Staat“ mit jeweils ca. 100 000 Wortformen. Alignment erfolgt auf unterschiedlichen Komplexitätsstufen: automatische Zuordnung von zwei oder mehreren äquivalenten Textfolgen, von Absätzen, von Sätzen (sentence-alignment) bis hin zur Zuordnung einzelner Satzteile zu ihrem Äquivalent in der anderen Sprache. Diese automatischen Zuordnungen beruhen prinzipiell auf einer detaillierten Kodifizierung von Satzgliedern, Satz- und Zeilenenden, Absätzen, Sequenzfolgen und Kapiteln auf der Basis der international üblichen SGML-Kodierung. Am Beispiel „Plato-Korpus“ soll dies nur andeutungsweise illustriert werden. Abbildung 1 zeigt, wie der Output einer automatischen Zuordnung für eine ganze Textfolge aussieht. Die Kopfzeilen mit bibliographischen Angaben wie Ausgabe, Kapitel, Sprache etc. und identische Numerierungen (z.B. n=327) gehören zu den wesentlichen Komponenten, die diese automatische Zuordnung möglich machen.

Abbildung 2 stellt dar, wie eine Parallelisierung einzelner Sätze aussehen kann. In diesem Fall ist sogar eine 1:1-Zuordnung der Sätze möglich, da der Dialog deutlich strukturiert ist.

Um zu relevanten Belegen zu kommen, wurde für den deutschen Plato-Text eine erste automatische Kollokationsrecherche mit einem IDS-eigenen Programm durchgeführt. Mit dem Einsatz dieses Tools erhält man die Ausgabe einer vollständigen Liste aller in einem Text/Korpus vorkommenden Zweiwortkombinationen unterschiedlichen Typs mit genauen Frequenzangaben. Einen kleinen Ausschnitt aus dieser Recherche stellt die folgende Liste (Abb. 3) dar, wobei sowohl die Häufigkeit der Wortverbindung (1. Zahl) als auch die Vorkommenshäufigkeit ihrer Einzelelemente belegt ist (Beispiel: *ganz richtig* tritt 59mal auf, das Adjektiv *ganz* insgesamt 140mal, das Adjektiv *richtig* insgesamt 118mal).

Eine solche Auflistung wird ebenso für das Französische durchzuführen sein. Kernpunkt all dieser Abfrageprogramme ist die statistische Erkennung jener Wortverbindungen, die häufiger als andere auftreten. Mit der IDS-Suchprozedur können auch ganz spezielle Wortart-Wortart-Kombinationen mit ihren einzelnen Vorkommenshäufigkeiten herausgefiltert werden. Für den Testfall „Plato“ wurden dazu in einem weiteren Schritt alle Adjektiv-Substantiv-Verbindungen in lemmatisierter Form isoliert. Einen Ausschnitt aus dieser Liste ist in Abb. 4 dargestellt.

668.57	Ganz richtig 59 140 118	390.48	Das ist 75 1102 286
644.01	wenn er 141 1101 740	385.39	müssen wir 61 1009 156
554.95	wenn sie 139 1438 740	384.63	So ist 66 1102 192
554.59	in der 184 2113 1104	349.33	glaube ich 48 753 106
540.37	ist es 137 966 1102	346.50	mit Recht 44 92 603
536.04	mein Sokrates 48 72 207	340.71	Ich glaube 36 106 187
520.32	des Staates 53 67 514	340.44	fragte ich 42 753 69
507.31	wollen wir 76 1009 170	338.78	der Seele 74 211 2113
494.51	meinst du 52 691 61	336.91	hast du 45 691 98
476.22	mit dem 90 589 603	333.35	denke ich 38 753 51
472.15	deiner Meinung 37 108 53	327.90	muß man 46 396 187
441.17	ich fort 45 48 753	319.75	Bei Zeus 24 65 40
402.41	fuhr ich 40 753 41	311.75	Sehr richtig 29 140 65
392.74	mein Freund 37 65 207		

Abbildung 3: Lemmatisierte Liste von Zweiwortverbindungen

298.26	breit Masse 19 22 41	61.45	früh Darlegung 5 37 17
133.08	wahr Sein 13 126 37	60.23	edl Haltung 5 33 21
122.66	unmittelbar Wiedergabe 7 10 12	56.83	ganz Staat 16 309 453
114.76	gleich Weise 12 108 52	53.33	jung Leute 6 34 109
111.79	voll Recht 12 67 93	52.00	gewohnt Lebensweise 3 6 7
105.63	hoch Gut 12 84 94	48.06	mittelbar Wiedergabe 3 6 12
85.70	tyrannisch Mensch 10 25 329	47.13	demokratisch Mensch 6 20 329
76.35	richtig Vorstellung 8 232 18	46.87	verschieden Natur 5 25 98
73.44	hoch Grad 5 84 5	46.52	wichtig Ding 5 23 111
72.70	öffentlich Leben 7 16 150	45.79	oligarchisch Mensch 6 22 329
71.23	musisch Erziehung 6 15 65	43.66	übrig Bürger 5 60 54

Abbildung 4: Lemmatisierte Frequenzliste von ADJ-SUBST-Kombinationen

Anschließend wurden nach acht der am häufigsten belegten bzw. auffälligsten Kollokationen im deutschen Text gesucht (*breit Masse*; *wahr Sein*, *gleich Weise*, *voll Recht*, *öffentlich Leben*, *edl Haltung*, *oligarchisch Mensch*, *ganz Seele*). Damit konnten komplette Beleglisten mit den jeweiligen Kollokationen in ihrem unmittelbaren Kontext (ca. 2-3 Sätze davor und danach) – vergleichbar auch mit den KWIC-Listen in COSMAS – erstellt und zugleich eine automatische Zuordnung des entsprechenden französischen Absatzes vorgenommen werden. Das heißt nicht in jedem Fall, daß auch im französischen Absatz eine Äquivalenzkollokation vorkommen mußte. Es bedeutet nur, daß beide Absätze die Übersetzung ein- und desselben Originaltextabschnitts darstellen. Bis auf wenige Fehlbelege gab es aber immer vergleichbare und eindeutig rekonstruierbare Äquivalente, wenn auch in ganz verschiedenen Explizierungen und Komplexitätsgraden.

Die folgenden drei Beispiele in Abb. 5 stellen dieses Programm mit der entsprechenden Suchanfrage [„Suche alle Belege für *breit Masse* in lemmatisierter Form und ordne den entsprechend analog numerierten französischen Absatz zu“] für einige ausgewählte Belege der Kollokation *breit Masse* und die entsprechenden Ergebnisse dar. Die Kollokation *breit Masse* im Deutschen und parallele Explizierungen im Französischen sind grafisch hervorgehoben. Zunächst ist festzustellen, daß die französische Übersetzung verschiedenartige sprachliche Ausdrücke und Explizierungen an jenen Stellen anbietet, an denen die deutsche Übersetzung als einzige

Variante *breit Masse* verwendet (*la masse, la masse du peuple, la plupart des gens, la masse des gens*).

*** Platon Deutsch - Französisch ***

— Suchergebnisse —

Sucheingabe : breit.(*(.??)) Masse

Gefunden: 19 Treffer

Beleg Nr. 1:

Fundstelle: Abschnitt 387. Paragraph b

Deutsch:

Homer und die anderen Dichter mögen uns, so bitten wir, doch nicht gram sein, wenn wir diese und ähnliche Stellen streichen; wir tun's nicht, weil die Verse undichterisch wären oder unangenehm zu hören *für die breite Masse*; gerade weil sie dichterisch sind, dürfen sie um so weniger von Kindern und Menschen gehört werden, die frei sein und die Knechtschaft mehr als den Tod fürchten sollen."

Französisch:

[...] nous priérons Homère et les autres poètes de ne pas se fâcher que nous les raturions, non pas comme non poétiques ou désagréables à entendre *par la masse*, mais parce que plus ils sont poétiques, moins il faut les faire entendre aux enfants et aux hommes dont on veut qu'ils soient des hommes libres, plus effrayés par l'esclavage que par la mort.

Beleg Nr. 5:

Fundstelle: Abschnitt 458, Paragraph d

Deutsch:

Oder hältst du das nicht für eine zwangsläufige Folge?"

„Nicht von mathematischer, sondern von erotischer Notwendigkeit, die noch viel zwingender *die breite Masse* überzeugt und mit sich zieht."

Französisch:

A ton avis, ce que je dis là n'est-il pas nécessaire ? – D'une nécessité qui n'est sans doute pas géométrique, dit-il, mais érotique, oui, nécessité qui risque d'être plus stimulante que la première pour convaincre et entraîner *la masse du peuple*. [...]

Beleg Nr. 18:

Fundstelle: Abschnitt 608, Paragraph a

Deutsch:

solange sie aber sich nicht verteidigen kann, werden wir ihr nie zuhören, ohne uns selbst durch unsere Überlegung und unseren Zaubergesang gefeit zu machen, aus Angst, wieder in die Liebe unserer Kindheit, *die Liebe der breiten Masse* zurückzufallen. Wir fühlen es, wie wenig ernst man eine solche Dichtung nehmen darf, die sich für wahr und ernsthaft gibt.

Französisch:

[...] mais tant qu'elle ne sera pas capable de présenter sa justification, nous l'écouterons sans cesser de nous chanter cet argument que nous disons, cette incantation, et en prenant garde de ne pas retomber dans notre amour d'enfance qui est aussi celui *de la plupart des gens*. Nous chanterons qu'il ne faut pas traiter avec sérieux une telle création poétique, comme si elle touchait à la vérité et était sérieuse, [...]

Abbildung 5

Diese Beleglisten bieten nun bereits in diesem Versuchsstadium Möglichkeiten zu einigen analytischen Beobachtungen. Noch ist die sprachliche Kompetenz des Linguisten gefragt, um die entsprechenden Parallelvarianten im französischen Text zu bestimmen. Inwieweit diese Zuordnungen ganz bzw. teilautomatisierbar sein werden, wird die Zukunft zeigen müssen. Für unser eigentliches Korpus können diese Kontextangaben dazu dienen, die jeweiligen „Einbettungsstrukturen“ zu analysieren, das heißt, die jeweiligen syntaktischen und lexikalisch-semantischen Umgebungen unterhalb der Satzgrenze, aber auch darüber hinausgehend (Satzfolgen und Textabschnitte), zu charakterisieren, um möglicherweise zu bestimmten kontextuellen Regularitäten bzw. wiederkehrenden Einbettungen signifikanter Kollokationen zu kommen. Bei zunehmender Komplexität wird dann auch die illokutive Umgebung immer relevanter (vgl. 2.4).

In (6) wird angedeutet, wie eine statistische Auswertung mit Häufigkeitsangaben und den dazugehörigen Textstellen aussehen könnte. Diese statistischen Angaben sind ein erster Versuch und können mit Hilfe weiterentwickelter Computerprogramme zunehmend verfeinert werden. Dabei kann zum einen auf bereits vorhandene Tools zurückgegriffen werden – sowohl externe als auch IDS-interne Entwicklungen –, zum anderen sind neue Entwicklungen für die speziellen Bedürfnisse des Projektes notwendig. Solche Übersichten sind vor allem bei umfangreichen Beleglisten aus großen Korpora von einem bestimmten Wert, da sie schnell eine Vorsortierung nach Frequenzen vornehmen. Konkordanzen mit Aussagen darüber, welche Varianten einer bestimmten Kollokation im anderssprachigen Text mit welcher Häufigkeit auftreten, helfen dem Linguisten, zunächst einmal einen Überblick über alle vorhandenen Belege zu gewinnen und eine erste Hierarchisierung der Bedeutsamkeit einzelner Beispiele vorzunehmen:

1. „breit* Masse“

Gesamtanzahl: 19 (2 Fehlbelege Nr. 3, 10)

a. *breit Masse - la masse*

Beleganzahl: 11

Beleg-Nr.: 1, 2, 4, 6, 7, 8, 9, 13, 14, 15, 17

b. *der breiten Masse - de la masse des gens*

Beleganzahl: 1

Beleg-Nr.: 0

c. *die breite Masse - la masse du peuple*

Beleganzahl: 1

Beleg-Nr. 5

d. *breit Masse - la plupart des gens* Beleganzahl: 3

Beleg-Nr.: 11, 16, 18

2. „wahr* Sein“

Gesamtanzahl der Belege: 13 (2 Fehlbelege Nr.: 2, 3)

a. *wahr Sein – qui est réellement*

Beleganzahl: 10 Belegnummer: 0, 1, 4, 5, 6, 7, 8, 9, 10, 12

b. *wahres Sein – l' être*

Beleganzahl: 1

Belegnummer: 13

Abbildung 6: Statistische Belegliste

Besonders ergiebig für die künftige Analyse ist ein Wörterbuchvergleich. Nachdem man eine Liste häufig vorkommender Kollokationen und entsprechender Übersetzungsäquivalente erstellt hat, scheint es geradezu geboten, traditionelle Wörterbücher heranzuziehen, um zu sehen, inwieweit die isolierten sprachlichen Formen bereits kodifiziert, als Wörterbucheintrag vorliegen bzw. inwieweit das Wörterbuch relevante Kontextuierungen anbietet. Damit kann aufgrund der Frequenzanalyse aber auch die Auswahl des Lexikographen in Frage gestellt bzw. ergänzt werden, wenn wesentliche Entsprechungen nicht im Wörterbuch auffindbar sind:

„breit* Masse“

Im Korpus vorkommende Varianten

la masse

la masse des gens

la masse du peuple

la plupart des gens

Vergleich mit Wörterbucheinträgen

LANGENSCHIEDT: *die Masse – la masse*
(Sachs/Villatte) *Volkmenge – la masse, la foule*
die breite Masse versteht davon nichts –
le gros public n'y comprend rien

PONS: *die Masse – la masse, (bes. Menschen) la foule*
(Weis/Mattutat) *die breite od. große Masse –*
la grande masse, les masses, le plus grand nombre

„wahr Sein“

Im Korpus vorkommende Varianten

qui est réellement

l' être

Vergleich mit Wörterbucheinträgen

LANGENSCHIEDT: *das Sein – l' être, (Wesenheit) – l' essence, (Dasein) – l' existence*
(Sachs/Villatte)

PONS: *das Sein – l' être, (Wesenheit) – l' essence, (Dasein) – l' existence*
(Weis/Mattutat)

„voll Recht“

Im Korpus vorkommende Varianten

*Selon toute apparence
le plus correctement
tout à fait exact
justement
nous aurons raison
tu disais vrai*

Vergleich mit Wörterbucheinträgen

LANGENSCHIEDT:	<i>das Recht – (begründeter Anspruch) le droit, la raison</i>
(Sachs/Villatte)	<i>mit Recht – avec raison, comme de juste, mit vollem od. gutem Recht – à bon droit, à juste titre, das Recht – le droit, (Berechtigung) la raison</i>
PONS:	
(Weis/Mattutat)	<i>mit Recht – avec raison, de droit, à bon droit, à juste titre mit vollem Recht – de plein droit, par droit et raison, à juste titre</i>

Abbildung 7

Bezogen auf unser Fallbeispiel seien nur einige Auffälligkeiten genannt: So entspricht die deutsche Kollokation *wahres Sein* – von einer einzigen Ausnahme im französischen Text abgesehen – der (alltagssprachlicheren) Verbalphrase *qui est réellement*. Interessant ist dabei, daß es durchaus auch französische Äquivalente gäbe, die dem philosophischen Charakter des Ausdrucks entsprächen (z.B. *l'être*). Es läßt sich nur vermuten, daß die Übertragung mit dem insgesamt stärker alltags-sprachlich geprägten Stil der französischen Übersetzung zusammenhängt, während der deutschen Fassung eine gehobenere Stilebene zuzuschreiben ist. Ein anderes Beispiel sind die Parallelformen von (*mit*) *vollem Recht*. Obwohl die zweisprachigen Wörterbücher sogar explizite Übersetzungsangebote für diese Kollokation anbieten (*à bon droit, à juste titre, de plein droit, par droit et raison*), werden diese Varianten nicht ein einziges Mal im französischen Text verwendet. Hier zeigt sich, daß man nicht nur die isolierte Fügung und ihr Äquivalent betrachten kann, sondern daß auch nicht selten der illokutiv-kommunikative Kontext ausschlaggebend ist. So ist die Äußerung „*Mit vollem Recht*“ gerade in (philosophischen) Dialogen auch als reine Bestätigungshandlung zu interpretieren. Für solche Bestätigungen kann es nun wieder ganz andere sprachliche Formen geben, die mit der eigentlichen Bedeutung der lexikalischen Einheit *Recht* nicht voll erfaßt wird (z.B. *tu disais vrai*).

Derzeit wird ein automatisches Verfahren (lexicon look-up) entwickelt, das es ermöglicht, ein Wörterbuch mit den jeweiligen Einträgen automatisch hinzuzuschalten. Danach kann man ohne jeglichen manuellen Aufwand sofort erkennen, welche Einzelwortäquivalente, Kombinationsmöglichkeiten und Kontextangebote bereits im jeweiligen (zweisprachigen) Wörterbuch existieren. Dadurch läßt sich bestimmen, welche Wortverbindungen bereits usueller Natur sind und zum lexikalischen Inventar einer Sprache gehören und welche eher okkasionellen Charakter besitzen bzw. vom Lexikographen nicht als usuell betrachtet wurden. Das Programm bietet auch die Möglichkeit, automatisch alle Belege, die einen entsprechenden Eintrag

im Wörterbuch besitzen, auszusondern und somit zugleich eine Liste der nicht-kodifizierten Äquivalente zu erhalten.

3. Zusammenfassung und Ausblick

Es scheint sinnvoll zu sein, zwei Analyseverfahren für das deutsch-französische Kollokationswörterbuch zu kombinieren:⁸

a) Automatische Erstellung und analytischer Vergleich von Kollokationslisten aus deutschen und französischen Texten des Parallelkorpus

Als Ergebnis werden vollständige Frequenzlisten aller Kombinationsmöglichkeiten von Zweiworteinheiten und ihrer kontextuellen Einbettungen vorliegen. Aus dieser Liste lassen sich diejenigen mit der höchsten Vorkommensfrequenz bzw. der entsprechenden statistischen Signifikanz bestimmen. Danach ist zu entscheiden, welche Kollokationsvorkommen relevant sind und welche eher marginal. Darüber hinaus wird es aber auch Kollokationen geben, die zwar nicht die quantitativen Kriterien erfüllen, denen aber aus ganz bestimmten Gründen dennoch eine Relevanz für den jeweiligen Text, die Textsorte oder die jeweilige Sprache generell zugesprochen werden muß. Dafür Relevanzkriterien zu entwickeln, scheint ein lohnendes Ziel für eine linguistische Analyse zu sein. Unter anderem muß man zu einer systematischeren Bestimmung von Kollokationskategorien im Spannungsfeld zwischen völlig freien Wortverbindungen einerseits und idiomatischen Wendungen andererseits kommen.

Als nächsten Schritt kann man dann die deutschen und französischen Befunde miteinander vergleichen, um herauszufinden, ob es invariante Kernbereiche von Kollokationen gibt, und wo die Variablen liegen. Es muß also gefragt werden, welche Kollokationen in beiden Sprachen ähnlich zentral sind, und welche z.B. im Deutschen charakteristischen Kollokationen im Französischen keine zentrale Rolle spielen (bzw. umgekehrt).

b) Automatische Erstellung und Systematisierung typischer Kontextuierungen von Kollokationen

Diese Vorgehensweise baut auf den in 2.3 beschriebenen Suchprozeduren und Parallelisierungen auf: Die als relevant bestimmten Kollokationen werden im jeweiligen Korpus recherchiert, mit ihren Kontexten und den Parallelvarianten in der anderen Sprache verglichen. Beispielsweise können alle Kontextbelege einer signifikanten Kollokation im Deutschen aufgelistet und gleichzeitig die äquivalenten französischen Übersetzungsbelege zugeordnet werden. Damit läßt sich zeigen, auf welcher unterschiedlichen Art und Weise ein und dieselbe Ausgangseinheit in der Quellsprache in den Übersetzungen in welchen Kontexten expliziert wird.

Im Mittelpunkt steht dabei immer die Frage, ob es Übersetzungsäquivalente gibt, die das klassische Wörterbuch nicht anbietet, und wenn ja, ob diese nur den Status von einzeltext- bzw. einzelsprecherbezogenen Ad-hoc-Bildungen haben, oder ob sie bereits Usus sind. Man könnte sie dann auch als Standardverwendungen bezeichnen, was sich nur auf der Basis großer Textmengen wirklich verifizieren läßt. Statistische

⁸ Zum methodischen Herangehen vgl. u.a. auch Windisch (1996).

Voruntersuchungen haben ergeben, daß ein großer Teil von Übersetzungseinheiten in solchen Parallelkorpora nicht im zweisprachigen Wörterbuch erfaßt ist (vgl. Geyken 1997). Es geht also vorrangig um die Darstellung von Kollokationen, die auf Grund ihrer Vorkommenshäufigkeit oder ihrer textuellen Relevanz durchaus signifikant sind, aber mit bisherigen lexikographischen Hilfsmitteln nicht erfaßt und beschrieben wurden. Solche Abweichungen, die ja möglicherweise gar keine Abweichungen sind, sondern ebenso Regularitäten und Standardvarianten darstellen, können nur mit Hilfe der vorgeschlagenen listen- und regelbasierten Recherchen und Analysen erfaßt werden.

Durch kontextuelle Gegenüberstellungen läßt sich weiterhin analysieren, inwieweit sich auch die entsprechenden Einbettungsstrukturen wiederholen. Man kann fragen: Gibt es spezifische sprachstrukturelle und kommunikative Umgebungen für bestimmte Kollokationen? Sind diese Umgebungen generalisierbar für bestimmte textuelle Konstellationen, Textsorten, Stilebenen usw.? Damit sind dann auch ganz praktische Bedürfnisse der Übersetzungsarbeit berührt: Das elektronische Kollokationswörterbuch wird nicht nur eine systematisierte Ordnung häufig verwendeter Kollokationen und ihrer Übersetzungsäquivalente enthalten, sondern auch *abrufbare, systematisierte Kontextmuster*.

Literaturverzeichnis

- Bárdosi, Vilmos/Ettinger, Stefan/Stölting, Cécile (1992): Redewendungen Französisch-Deutsch. Thematisches Wörter- und Übungsbuch. Tübingen.
- Bahns, Jens (1996): Kollokationen als lexikographisches Problem. Eine Analyse allgemeiner und spezieller Lernerwörterbücher des Englischen. (= Lexicographica: Series Maior 74). Tübingen.
- Boguraev, Branimir/Pustejovsky, James (1996): Corpus Processing for Lexical Acquisition. (= Language, Speech, and Communication). Massachusetts.
- Burger, Harald/Buhofer, Annelies/Sialm, Ambros (1982): Handbuch der Phraseologie. Berlin/New York.
- Fleischer, Wolfgang (1982): Phraseologie der deutschen Gegenwartssprache. Leipzig.
- Geyken, Alexander (1997): Matching Corpus Translations with Dictionary Sense: Two Case Studies. In: Teubert, W. (Hg.): International Journal of Corpus Linguistics. Mannheim. (erscheint).
- Hausmann, Franz Josef (1985): Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: Bergenholtz, H./Mugdan, J. (Hg.): Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984. (= Lexicographica 3). Tübingen, S. 118-129.
- Hausmann, Franz Josef (1989): Le dictionnaire de collocations. In: Hausmann, F.J./Reichmann, O./Wiegand, H.E./Zgusta, L. (Hg.): Wörterbücher. Ein internationales Handbuch zur Lexikographie. Erster Teilband. (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1), S. 1010-1019.
- Hausmann, Franz Josef (1993): Was ist und was soll ein Kontextwörterbuch? (Einführung). In: Ilgenfritz, P./Stephan-Gabinel, N./Schneider, G.: Langenscheidts Kontextwörterbuch Französisch-Deutsch. Ein neues Wörterbuch zum Schreiben, Lernen, Formulieren. Berlin/München/Wien/Zürich/New York.

- Ilgenfritz, Peter/Stephan-Gabinel, Nicole/Schneider, Gertraud (1993): Langenscheidts Kontextwörterbuch Französisch-Deutsch. Ein neues Wörterbuch zum Schreiben, Lernen, Formulieren. Mit einer Einführung von F. J. Hausmann, Berlin/München/Wien/Zürich/New York.
- Korhonen, Jarmo (1992): Idiome als Lexikoneinheiten. Eine Auswahl von Beschreibungsproblemen. In: Korhonen, J.(Hg.): Phraseologie und Wortbildung – Aspekte der Lexikonerweiterung. Finnisch-deutsche sprachwissenschaftliche Konferenz, 5.-6. Dezember 1990 in Berlin. (= LA 284). Tübingen. S. 1-20.
- Lehr, Andrea (1996): Kollokationen und maschinenlesbare Korpora. Ein operationales Analysemodell zum Aufbau lexikalischer Netze. (= Reihe Germanistische Linguistik 168). Tübingen.
- Koller, Werner (1992): Einführung in die Übersetzungswissenschaft. 4., völlig neu bearb. Aufl. Heidelberg/Wiesbaden.
- Kleiber, Georges (1993): Prototypensemantik. Eine Einführung. Übersetzt v. Michael Schreiber. Tübingen.
- Palm, Christine (1995): Phraseologie. Eine Einführung. Tübingen.
- Sachs, Karl/Villatte, Césaire (1994): Langenscheidts Großwörterbuch Französisch. Teil I. Französisch-Deutsch. Hg. v. E. Weis. 8. Aufl. der V. Neubearb. 1979. Berlin/München/Wien/Zürich.
- Sachs, Karl/Villatte, Césaire (1979): Langenscheidts Großwörterbuch Französisch. Teil II. Deutsch-Französisch. Völlige Neubearb. Hg. v. W. Gottschalk u. G. Bentot. 13. Aufl. der V. Neubearb. 1968. Mit Nachtrag 1979. Berlin/München/Wien/Zürich.
- Sandig, Barbara (Hg.) (1994): EUROPHRAS 92. Tendenzen der Phraseologieforschung. (= Studien zur Phraseologie und Parömiologie 1). Bochum.
- Schemann, Hans/Raymond, Alain (1994): PONS Idiomatik Deutsch-Französisch. Dictionnaire Idiomatique Allemand-Français. Stuttgart/Dresden.
- Sinclair, John (1991): Corpus, Concordance, Collocation. Oxford
- Stolze, Radegundis (1994): Übersetzungstheorien. Eine Einführung. Tübingen.
- Stubbs, Michael (1996): Text and Corpus Analysis. Computer-assisted Studies of Language and Culture. (= language in society 23). Oxford.
- Teubert, Wolfgang (1996a): Comparable or Parallel Corpora? In: International Journal of Lexicography, Vol. 9 No.3. Oxford, S. 238-276.
- Teubert, Wolfgang (1996b): Deutsch-französische Verständigung. Ein Übersetzungswerkzeug für das 21. Jahrhundert. In: Sprachreport 2/96, S. 7-11.
- Weis, Erich (1993): PONS Großwörterbuch Französisch-Deutsch/Deutsch-Französisch. Unter Mitwirkung v. H. Mattutat. Neubearb. unter Mitwirkung v. C. Nugue. Erw. Neubearb. 1. Aufl. v. 1988 (Nachdruck). Stuttgart/Dresden.
- Wilss, W. (1977): Übersetzungswissenschaft. Probleme und Methoden. Stuttgart.
- Windisch, Eric (1996): Projektstudie Korpus-basierte Kollokationswörterbücher. 1. Band: Adjektive und ihre Bezugsubstantive. Mannheim.
- Zimmer, Rudolf (1990): Äquivalenzen zwischen FRANZÖSISCH und DEUTSCH. Theorie – Korpus – Indizes. Ein Kontextwörterbuch. Tübingen.

Dr. Kathrin Steyer
 Dr. Wolfgang Teubert
 Institut für deutsche Sprache
 R 5, 6-13
 D-68016 Mannheim
 E-mail: steyer@ids-mannheim.de
 wteubert@ids-mannheim.de